On the System Optimum of Traffic Assignment in $M/G/c/c$

State-dependent Queueing Networks

F. R. B. Cruz∗† T. van Woensel‡ J. MacGregor Smith§ K. Lieckens¶

fcruz@est.ufmg.br T.v.Woensel@tm.tue.nl jmsmith@ecs.umass.edu Kris.Lieckens@ua.ac.be

February 23, 2009

Abstract — The classical Wardrop System Optimum assignment model assumes that the users will cooperate with each other in order to minimize the overall travel costs. The importance of the System Optimum model lies on its well-recogznized ability of producing solutions that correspond to the most efficient way of using the scarce resources represented by the street and road capacities. In this paper, we present a version of the System Optimum model in which the travel costs incurred on each path come from $M/G/c/c$ state-dependent queueing networks, a stochastic travel time estimation formula which takes into account congestion effects. A differential evolution algorithm is proposed to solve the model. We motivate this version of the problem in several ways and computational results show that the proposed approach is efficient.

Keywords — System Optimum; traffic assignment; queueing networks; genetic algorithms.

1 INTRODUCTION

There have been successful attempts in the literature to model how users select their route in a congested network (for instance, see Helbing et al., 2005, and references therein). Two major streams of work can be distinguished: the System Optimum (SO) models versus the User Equilibrium (UE) models.

The User Equilibrium model, assuming perfect knowledge of the travel costs, states that drivers will choose the best route according to Wardrop’s first principle. This principle is equivalent to a mixed-strategy Nash equilibrium of an $n$-player, non-cooperative game (Bell & Cassir, 2002). The Deterministic User Equilibrium (UE) is an important classical traffic assignment model approach (Sheffi, 1985), which even recently keeps receiving improvements (see, for instance, the recent paper by Watling, 2006). In equilibrium, routes carrying a positive flow will have equal travel costs. The disadvantage of the User Equilibrium model is that the scarce resources (street and road capacity) may be used in an inefficient way (Helbing et al., 2005).

In contrast, the classical Wardrop System Optimum (SO) assignment model, assumes that all users are able to cooperate with each other in order to minimize the overall system-wide travel costs (Sheffi, 1985). Even though the System Optimum (SO) assignment model is based on a rather non-realistic behavioral assumption, we argue that its solution may be seen as a result of a well-succeeded control action on the transportation network, such as, for instance, by route inducement (Moreno-Quintero, 2006). In other words, signal timings may be re-optimized and alternative routes may be re-defined in response to an increase in demand. It is well-known that traffic lights and adaptive routing can improve the flow (Poli Jr. & Monteiro, 2005), depending on the traffic densities (e.g. using DRIPS, Dynamic Routing Information Panel Systems). Next to this, several paradoxes show the deficiency of the UE optimum compared to the SO model. For example, Braess’s paradox shows that adding extra capacity to a network, when people selfishly choose their own route, can reduce overall performance (Braess, 1968; Braess et al., 2005). A similar result has been observed by Sheffi & Daganzo (1978). On the other hand, Charnes & Klingman (1971) showed that both increasing supply and demand could counter-intuitively lead to a reduction in total costs. In any case, both paradoxes show that transport planners should not trust in the users’ selfish actions when optimizing the traffic network. As such,
these paradoxes enforce the necessity of the System Optimum (SO) assignment model.

One major problem in the above models is that the travel times are usually assumed to be either deterministic or approximate stochastic models. Typically, the SO models express the travel costs in terms of deterministic travel time functions (Prashker & Bekhor, 2000), yet these times are known to be rather variable between trips, within and between days. The relevant travel time models are usually built on the classical formulas that have been constructed over the past 40 years. For instance, the well-known BPR (Bureau of Public Roads, 1964) was developed in 1964 using data from the Highway Capacity Manual.

Kimber & Hollis (1979) developed another travel time formula based upon an approximation to the time-dependent $M/G/1/\infty$ model. Since analytical expressions for the transient $M/G/1/\infty$ model are intractable, they developed an approximation based upon a coordinate transformation technique to adjust the steady state formula to account for the transient effects of the queue. In their approach, they can account for existing traffic on the highway link, but they fix the service rate of the traffic link $\mu$, the queue is infinite in capacity, and there is only one server for the traffic. Subsequently Akcèlïk (1991) extended the work of Kimber & Hollis with formulas based upon the coordinate transformation technique that are recognized as efficient to model the travel times, especially under congestion during rush hours, when the demand far exceeds capacity (Ceylan & Bell, 2005). The performance of Akcèlïk’s model is similar to Kimber & Hollis. Under these ‘typical’ link performance functions, good solution methods are well-known. We will argue that another stochastic approach is more powerful based on state dependent queues because it can also handle general service times, multiple servers, and has transient as well as steady-state solutions. It is a true stochastic approach with no approximations. Figure 1, presents results from many empirical studies for North American roads (Drake et al., 1967; Edie, 1961; Greenshields, 1935; Transportation Research Board, 2000; Underwood, 1961) under low traffic, the queueing approach is close to classical and accurate formulas, such as BPR and Akcèlïk’s, as seen in Fig. 2.

In particular, we introduce a stochastic version of the SO model in which the costs incurred on each path come from $M/G/c/c$ state-dependent queueing networks. This latter model is a stochastic travel time estimation formula that takes into account these important congestion effects. The $M/G/c/c$ state-dependent queueing models originated with the work of Yuhaski & Smith (1989) for pedestrian traffic flows. This paper formed the foundation of all the subsequent models used in this approach to the travel time flow modeling problem. Following this were the papers of Cheah & Smith (1994) which generalized the process and showed that the state dependent queue was quasi-reversible and Jain & Smith (1997) which showed how the state dependent queues could be used for modeling vehicular congestion. In Sec. 3.1.2, we will describe in detail the elaboration of the $M/G/c/c$ state-dependent queueing model. For a review on the use of queueing models to model traffic flows and congestion, the reader is referred to the paper by van Woensel & Vandaele (2007). Another successful attempt to refine the travel time estimation may be found in the paper by García-Ródenas et al. (2006).

Fig. 2 shows typical travel time functions (recently used, for instance, by Ghatee & Hashemi, 2009; Pursals & Garzón, 2009) in comparison with the $M/G/c/c$ state-dependent queueing model functions (Jain & Smith, 1997), for a 1-mile long freeway, with free-flow speed 62.5 mph (100 km/h), and capacity 2,400 veh/h, based on the Highway Capacity manual (Transportation Research Board, 2000). In addition, Fig. 3 shows how the travel time functions behave as a function of the arrival rate for several single links admitting an $M/G/c/c$ state-dependent queue to model the road traffic. Note that under low traffic, the queueing approach is close to classical and accurate formulas, such as BPR and Akcèlïk’s, as in Fig. 2.

Important to note from Fig. 2 and Fig. 3 is that the $M/G/c/c$ travel time function is not convex but $S$-shaped, which will produce many local optima. Consequently, the introduction of these stochastic $M/G/c/c$ state-dependent models will make the SO problem computationally more challenging as multiple solutions may be present. The Frank-Wolfe algorithm is a convex combination algorithm (Frank & Wolfe, 1956) that has been often used for determining the equilibrium flows in transportation networks. However, since our
System Optimum in $M/G/c/c$ Queues

Contributions

The main contributions of this paper are twofold. First, we propose a stochastic extension to the SO model by applying state-dependent $M/G/c/c$ queueing network models in order to estimate the travel times, usually the main factor for route selection (Prashker & Bekhor, 2000). The consequence of selecting these more realistic travel time functions is that the objective function in the System Optimum model exhibits multiple local optima.

Secondly, because of these local optima, we propose a different way to solve the SO by using a Differential Evolution (DE) heuristic, which is part of the family of Genetic Algorithms (GA). This hybrid modeling approach (finite queueing networks and Differential Evolution) results in efficient and acceptable solutions for the problem on-hand.

Besides, the following results are obtained in the paper.

1. Networks of $M/G/c/c$ state-dependent queues are an effective way of modeling travel times in traffic networks;
2. The Differential Evolution (DE) heuristic proves to be an efficient optimization tool for the problem on-hand;
3. The algorithms proposed provide fast and good-quality solutions for realistic and complex topologies.

Outline of Paper

This paper is structured as follows. In Sec. 2 the mathematical programming formulations for the classical traffic assignment models are presented in detail. Sec. 3 describes the algorithm for solving the SO, as well as the performance evaluation algorithm. Sec. 4 focuses on detailed computational experiments, some of them based on an actual urban traffic network. Finally, Sec. 5 summarizes the paper and discusses open questions for future research in the area.

## 2 Mathematical Programming Formulations

Well-known from the literature, the main equilibrium formulations are the System Optimum (SO) and the Deterministic User Equilibrium (UE), which are briefly reviewed as follows. These models can be classified according to the behavioral assumption governing route choice and, generally, the major factor for route choosing are the expected travel times (Prashker & Bekhor, 2000). The network notation used is summarized in Tab. 1.
Table 1: Basic network notation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathcal{N} )</td>
<td>node (index) set</td>
</tr>
<tr>
<td>( \mathcal{A} )</td>
<td>arc (index) set</td>
</tr>
<tr>
<td>( \mathcal{R} )</td>
<td>set of origin nodes; ( \mathcal{R} \subseteq \mathcal{N} )</td>
</tr>
<tr>
<td>( \mathcal{S} )</td>
<td>set of destination nodes; ( \mathcal{S} \subseteq \mathcal{N} )</td>
</tr>
<tr>
<td>( K_{rs} )</td>
<td>set of paths connecting origin-destination ((O-D)) pair ( r - s; r \in \mathcal{R}, s \in \mathcal{S} )</td>
</tr>
<tr>
<td>( x_a )</td>
<td>flow on arc ( a; x = (\ldots, x_a, \ldots) )</td>
</tr>
<tr>
<td>( c_a(x_a) )</td>
<td>travel time on arc ( a; c(x) = (\ldots, c_a(x_a), \ldots) )</td>
</tr>
<tr>
<td>( f_k^{rs} )</td>
<td>flow on path ( k ) connecting ((O-D)) pair ( r - s; f^{rs} = (\ldots, f_k^{rs}, \ldots); f = (\ldots, f^{rs}, \ldots) )</td>
</tr>
<tr>
<td>( c_k^{rs} )</td>
<td>travel time on path ( k ) connecting ((O-D)) pair ( r - s; c = (\ldots, c_k^{rs}, \ldots); c = (\ldots, c^{rs}, \ldots) )</td>
</tr>
<tr>
<td>( q^{rs} )</td>
<td>trip rate between origin ( r ) and destination ( s; q^{rs} = q^{rs} )</td>
</tr>
<tr>
<td>( \delta_{a,k} )</td>
<td>indicator variable: ( \delta_{a,k} = \left{ \begin{array}{ll} 1, &amp; \text{if link } a \text{ is on path } k \text{ between } O-D \text{ pair } r - s, \ 0, &amp; \text{otherwise}; \end{array} \right. )</td>
</tr>
<tr>
<td>( (\Delta^<em>)<em>{a,k} = \delta</em>{a,k}^</em>; \Delta = (\ldots, \Delta^*, \ldots) )</td>
<td></td>
</tr>
</tbody>
</table>

2.1 User Equilibrium (UE)

The User Equilibrium model results from a choice of the best route according to Wardrop’s first principle. In other words, no driver could possibly reduce the corresponding travel time by moving unilaterally to another route. Several mathematical formulations appear in the literature for the model. Beckmann et al. (1956) represented the UE model by a formulation assuming a link cost as a continuous increasing function of the link flows, as follows

\[
\min z(x) = \sum_a \int_0^{x_a} c_a(w)dw,
\]

subject to

\[
\sum_k f_k^{rs} = q^{rs}, \quad \forall \ r, s,
\]

\[
x_a = \sum_r \sum_s f_k^{rs} \delta_{a,k}^{rs}, \quad \forall \ a,
\]

\[
f_k^{rs} \geq 0, \quad \forall \ k, r, s,
\]

in which \( x_a \) is the flow on link \( a, c_a(w) \) is the travel cost on link \( a \) as a function of the flow \( w, f_k^{rs} \) is the flow on route \( k \) between origin \( r \) and destination \( s \), and \( q^{rs} \) is the demand between \( r \) and \( s \) (for the complete notation, see Tab. 1).

The solution of this model are the equilibrium route flows defined as

\[
f_k^{rs}(c_k^{rs} - g^{rs*}) = 0, \quad c_k^{rs} - g^{rs*} \geq 0, \quad \forall \ k, r, s,
\]

in which \( c_k^{rs} \) is the cost on route \( k \) between \( r \) and \( s \) and \( g^{rs*} \) is the user equilibrium route cost between \( r \) and \( s \). Note that all travel costs on each path carrying a positive flow are equal at the equilibrium.

2.2 System Optimum (SO)

The System Optimum formulation is equivalent to a situation in which users cooperate (or are forced to cooperate) with each other in order to minimize the total travel costs. According to Wardrop’s second principle, the SO model is formulated as follows

\[
\min z(x) = \sum_a x_a c_a(x_a),
\]

subject to:

\[
\sum_k f_k^{rs} = q^{rs}, \quad \forall \ r, s,
\]

\[
x_a = \sum_r \sum_s f_k^{rs} \delta_{a,k}^{rs}, \quad \forall \ a,
\]

\[
f_k^{rs} \geq 0, \quad \forall \ k, r, s.
\]

The optimum solution is reached when the marginal travel costs on each path carrying a positive flow are equal, that is

\[
f_k^{rs}(g_k^{rs} - g^{rs*}) = 0, \quad g_k^{rs} - g^{rs*} \geq 0, \quad \forall \ r, s,
\]

in which \( g_k^{rs} \) is the marginal cost on route \( k \) and \( g^{rs*} \) is the optimal marginal cost, both between \( r \) and \( s \).

2.3 Remarks

The relationship between the two mentioned classical models above has long been studied in the literature (see Frashker & Bekhor, 2000, and references therein). We cannot leave unnoted the comparison between UE and SO solutions, with its most popular example, the Braess’ paradox. Also, the SO problem can be written as a special UE model with an adjusted travel cost function. However, we only propose a solution algorithm.
for the SO model in this article. Again, the main goal is to improve the traffic network as a whole by minimizing the total travel costs.

Note that the travel costs, \( c_a(x_a) \), are usually expressed in terms of times (Prashker & Bekhor, 2000) and are usually given from classical formulas. In this paper, the aim is to investigate a different travel time estimation formula, namely an \( M/G/c/c \) state-dependent queueing network based formula. For classical and \( M/G/c/c \) based formulas, the reader is referred to Fig. 2, which shows the usual link performance functions in comparison with the function given by the \( M/G/c/c \) state-dependent queueing model, for a 1-mile long freeway, with free-flow speed 62.5 mph (100 km/h).

3 Algorithms Proposed

Concerning how people have solved the UE and SO models, many algorithms have been proposed. To cite a few, results have been reported with dual algorithms (Hearn & Lawphongpanich, 1990), parallel algorithms (Ho, 1990), and Lagrangian based algorithms (Larsson & Patriksson, 1995). Recently published results on stochastic algorithms for the solution of traffic assignment problems include the papers by Ceylan & Bell (2005) and Patriksson (2006).

The algorithm proposed in Fig. 4 is based on a simple and easy to implement scheme. Note that the algorithm solves iteratively the routing probabilities \( p_{ij} \) and the performance evaluation, toward the minimization of the objective function \( \sum_a x_ac_a(x_a) \). The main steps are detailed as follows.

<table>
<thead>
<tr>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>read graph, ( G(V, A) )</td>
</tr>
<tr>
<td>read arrival rates, ( \lambda_a ), ( \forall a \in V )</td>
</tr>
<tr>
<td>read link lengths, ( l_{ij} ), ( \forall (i, j) \equiv a \in A )</td>
</tr>
<tr>
<td>repeat</td>
</tr>
<tr>
<td>/* generate routing probabilities */</td>
</tr>
<tr>
<td>generate ( p_{ij} ), ( \forall (i, j) \equiv a \in A )</td>
</tr>
<tr>
<td>/* compute flows and travel times */</td>
</tr>
<tr>
<td>compute ( x_a, c_a(x_a) ), ( \forall a \in A )</td>
</tr>
<tr>
<td>/* compute objective function */</td>
</tr>
<tr>
<td>compute ( \sum_a x_ac_a(x_a) )</td>
</tr>
<tr>
<td>until convergence is reached</td>
</tr>
<tr>
<td>write ( p_{ij}^{(opt)} ), ( \forall a \in A )</td>
</tr>
<tr>
<td>end algorithm</td>
</tr>
</tbody>
</table>

![Figure 4: Algorithm for SO traffic assignment](image)

3.1 Computing Flows and Travel Times

Note that the optimization heuristic needs an estimate for the flows and travel times, \( x_a \) and \( c_a(x_a) \), which may be done as follows. First, the single link modeling is presented. Then, a model for an entire transportation network is described.

3.1.1 Single Link Modeling

The traffic movement area of a single link may be seen as \( c \) parallel servers to its occupants, which is also the total number of users allowed in a system that has no buffer or waiting space. Second, based on the empirical results presented in Fig. 1, the service time for the occupants depends on the number of users currently in the system. As a consequence, an \( M/G/c/c \) state-dependent queueing model seems to be a reasonable tool to describe a single link (Yuhaski & Smith, 1989).

The limiting probabilities for the random number of entities \( N \) in an \( M/G/c/c \) queueing model, \( p_n \equiv \Pr[N = n] \), are as follows (Yuhaski & Smith, 1989)

\[
p_n = \left\{ \frac{\lambda E[T_1]^n}{n!f(n)f(n - 1)\cdots f(2)f(1)} \right\} p_0,
\]

in which \( n = 1, 2, \ldots, c, p_0 \) is the empty system probability, given by

\[
p_0^{-1} = 1 + \sum_{i=1}^{c} \frac{\lambda E[T_1]^i}{i!f(1)f(2)\cdots f(i)}.
\]

\( \lambda \) is the arrival rate, \( E[T_1] = l/V_1 \) is the expected service time of a lone vehicle in the traffic space of length \( l \), considering that \( V_1 \) is the speed of a lone vehicle, and \( c \) is the capacity of the traffic space

\[
c = \lfloor klw \rfloor,
\]

in which \( \lfloor x \rfloor \) is the largest integer not superior to \( x \), \( l \) is the length, \( w \) is the width in number of lanes, and \( k \) is the capacity of the link per length-unit per lane. Considering vehicular related applications and realizing that \( k \) represents the jam density parameter (veh/mi-lane), normally it ranges from 185-265 veh/mi-lane.

Notice that in Eqs. (1) and (2), \( f(n) = V_0/V_1 \) is the service rate, considered to be the ratio of the average speed of \( n \) users in the link to that of a lone occupant \( V_1 \). Basically, what one wants is that the congestion model represents the effect depicted in Fig. 1, in which the service rate depends on the number of user in the system.

Successful in the past, presenting consistent and robust empirical results (Yuhaski & Smith, 1989), was an exponential model in which the service rate decays following the expression

\[
f(n) = \exp \left( -\frac{n - 1}{\beta} \right)^\gamma,
\]

with

\[
\gamma = \log \left( \frac{\log(V_0/V_1)}{\log(V_0/V_1)} \right) / \log \left( \frac{a - 1}{b - 1} \right),
\]

and

\[
\beta = \frac{a - 1}{\log(V_1/V_0)^{1/\gamma}} = \frac{b - 1}{\log(V_1/V_0)^{1/\gamma}}.
\]
The values $a$ and $b$ are arbitrary points used to adjust the exponential curve. In vehicular related applications, commonly used values are $a = 20$ veh/mi-lane and $b = 140$ veh/mi-lane corresponding to densities of 20 and 140 veh/mi-lane respectively. Looking at the curves presented in Fig. 1, reasonable values for such points are $V_a = 48$ mph and $V_b = 20$ mph.

From Eq. (1), important performance measures can be derived

$$
\begin{align*}
\rho_i &= \Pr[N = 0], \\
\lambda &= \lambda(1 - \rho_i), \\
L &= E[N] = \sum_{n=1}^{c}np_n, \\
W &= E[T] = L/\theta,
\end{align*}
$$

in which $\rho_i$ is the blocking probability, $\theta \equiv x_a$ is the throughput in veh/h, $L$ is the expected number of customers in the link (also known as work-in-process, WIP), and $W \equiv c_a(x_a)$, here derived from Little’s formula, is the expected service time in hours.

3.1.2 Queueing Network Modeling

Deriving performance measures for $M/G/c/c$ state-dependent queues configured in networks is a task considerably more complex because of the routing probabilities that will define the input in each queues and because of the inter-blocking effects. An algorithm available is the Generalized Expansion Method (GEM), successfully used in the past to estimate performance measures for finite queueing networks.

Well described in many papers, in particular in the recently published paper by Kerbache & Smith (2000), the GEM consists of creating for each finite queue, representatives of customers, and $WIP$, and $W \equiv c_a(x_a)$, here derived from Little’s formula, is the expected service time in hours.

The values $\rho_i$ is the blocking probability, $\theta \equiv x_a$ is the throughput in veh/h, $L$ is the expected number of customers in the link (also known as work-in-process, WIP), and $W \equiv c_a(x_a)$, here derived from Little’s formula, is the expected service time in hours.

The iterative algorithm presented in Fig. 6 was recently proposed by Cruz & Smith (2007). The algorithm has been used in the context of vehicular traffic network modeling for the first time here in this paper. Firstly, a pre-evaluation is performed, Fig. 6-a. The performance evaluation algorithm chooses an arbitrary node, $j$, from set $V$ but not from set $Q$ (in which $Q$ is the set of nodes already evaluated), such that for all arc $(i, j) \in A$, vertex $i$ has been evaluated already. Then, vertex $j$ has computed its blocking probability $p_k^{(j)}$ and its arrival rate, from

$$
\theta_j = \lambda_j \times (1 - p_k^{(j)}).
$$

These service rates are then forwarded as arrival rates to the downstream nodes (if they exist), and vertex $j$ is included in set $Q$. Notice that the pre-evaluation step is a variant of Dijkstra’s minimum path algorithm (Dijkstra, 1959).

The GEM includes also an evaluation step, Fig. 6-b. This second part of the algorithm seeks flow conservation, that is

$$
\theta_j \leq \lambda_j + \sum_{i \in \{i, j\} \in A} \theta_i p_{kj}, \forall j \in V.
$$

The evaluation algorithm is a Dijkstra’s labeling algorithm working in reverse. Notice that the performance evaluation algorithm must have available the routing probabilities $p_{ij}$ before it can compute all the performance measures.

3.2 Generating routing probabilities

For the routing probabilities, $p_{ij}$, we propose to use the Differential Evolution (DE) heuristic, which is part of the broader family of Genetic Algorithms (GA). The following characteristics of the DE algorithm, valid for continuous space optimization problems (Storn & Price, 1997), justify its use as an appropriate solution method for our problem:

- it is simple, fast and robust;
- it has a superior global optimization ability;
So the mutation that improves the object vectors reflects
mines the distribution of the object vector differences.
response of these object vectors to the objective function
ence of randomly sampled pairs of object vectors. The
defined distribution function, while DE uses the differ-
context of EA’s, mutation is based on the output of a pre-
membership algorithm based on populations. The DE dif-
fittest. It is basically a computerized search and opti-
mixed-discrete programming, simulated annealing, genetic algorithm,
methods (branch & bound using sequential quadratic
end algorithm

Figure 6: Performance evaluation algorithm

- it can easily be implemented in a parallel computing
environment, which speeds up the optimization;
- it is effective in nonlinear optimization and can be very
  easily adapted for mixed parameter optimization;
- it does not require a differentiable objective func-
tion;
- it operates on flat surfaces;
- it can provide multiple solutions in a single run.

In addition, Babu & Sastry (1999) found the technique of DE to be the best evolutionary computation method after the study of seven difficult design and control MINLP problems in chemical engineering. For hard non-linear objective functions with multiple non-trivial constraints, Lampinen & Zelinka (1999) report solutions found by the DE that outperform any of the competing methods (branch & bound using sequential quadratic programming, integer-discrete-continuous non-linear programming, simulated annealing, genetic algorithm, non-linear mixed-discrete programming, …).

The DE is an improved version of genetic algorithms, which belongs to the class of evolutionary algorithms (EA) that are based on the principle of survival of the fittest. It is basically a computerized search and optimization algorithm based on populations. The DE differs from EA’s in the way the mutation is driven. In the context of EA’s, mutation is based on the output of a predefined distribution function, while DE uses the difference of randomly sampled pairs of object vectors. The response of these object vectors to the objective function determines their distribution, which on its turn determines the distribution of the object vector differences. So the mutation that improves the object vectors reflects information of the objective function it is optimizing. Instead of using only local information for each object vector, the DE mutates all object vectors with the same universal distribution. In this way the whole search space is covered and a global optimum can be found.

The method is defined as a parallel direct search method which operates on a population $P_G$ of constant size that is associated with each generation $G$ and consists of $NP$ vectors, or candidate solutions, $X_{p,G}$, $p = 1, 2, \ldots, NP$. Each vector $X_{p,G}$ consists of $D$ decision variables $X_{o,p,G}$, $o = 1, 2, \ldots, D$. This is briefly summarized as:

$$P_G = \{X_{1,G}, X_{2,G}, \ldots, X_{p,G}, \ldots, X_{NP,G}\}$$

$$X_{p,G} = \{X_{1,p,G}, X_{2,p,G}, \ldots, X_{o,p,G}, \ldots, X_{D,p,G}\}$$

$$G = 1, \ldots, G_{\text{max}}$$

$$NP \geq 4$$

Each routing probability is then considered as the decision variable, $p_{ij} \equiv X_{o,p,G}$.

The different steps of the algorithm are:

**Step 1:** Choose a strategy. Price & Storn (2006) suggested ten different strategies of DE, i.e. the DE/rand/1/bin, DE/rand/2/bin, and DE/current-to-rand/1 schemes.

**Step 2:** Initialize the key parameters of control. The user-defined control parameters, which remain constant during the search process, are the crossover constant $CR$, the population size $NP$, the mutation scaling factor $F$, the coefficient of combination $K$ and the maximum number of generations $G_{\text{max}}$. 

```
algorithm initialize set of labeled nodes, P ← ø
initialize maximum throughput, $\bar{\eta}_{\text{max}}$ ← ∞, ∀i ∈ V
while P ≠ V
  choose i such that (i ∈ V) and (i ∉ P)
  if {j : (i, j) ∈ A} ⊆ P then
    /* update performance measures */
    $E[T_i] ← \min E[T_i]$, s.t: $\theta_i ≤ \bar{\eta}_{\text{max}}$
    $E[T_i] ≥ I_i/V_i$
    compute $Pr[N = c_i]$, $\theta_i$, $E[T_i]$
    $/*$ backward propagate to predecessors */
    for all $k ∈ \{k' | (k', i) ∈ A\}$ then
      $\lambda_k ← \lambda_k + \theta_k p_{jk}$
  end if
end while
```

```
algorithm get routing probabilities, $p_{ij}$, ∀ (i, j) ∈ A
initialize set of labeled nodes, P ← ø
while P ≠ V
  choose j such that (j ∈ V) and (j ∉ P)
  if {i : (i, j) ∈ A} ⊆ P then
    /* compute performance measures */
    $E[T_i] ← I_i/V_i$
    compute $Pr[N = c_i]$
    compute $\theta_j$
    compute $L_i, W_j$
    /* forward information to successors */
    for all $k ∈ \{k' | (j, k') ∈ A\}$ then
      $\lambda_k ← \lambda_k + \theta_j p_{jk}$
    end for
  end if
end while
```


**Step 3:** Initialize the population. The initial population \( P_G = 0 \) provides us with a starting solution for optimum seeking and is chosen randomly within the bounds of the parameters that are set by the constraints. It should cover the entire variable space.

**Step 4:** Evaluate the profit of each vector and find the one with the highest profit. The objective function have to be evaluated for each vector in the population, after which the best one can easily be determined.

**Step 5:** Perform mutation and recombination. Mutation aims to keep a population robust and to search a new area. Mutation involves adding a randomly generated step to one or more parameters of an existing object vector in order to move existing object vectors in the right direction by the right amount at the right time. DE mutates an object vector by adding the weighted difference of randomly sampled pairs of vectors in the current population \( P_G \). The mutated vector that will be used to build the population of the next generation is denoted by \( V_{p,G+1} \). Recombination, or crossover, is complementary to mutation and builds trial vectors out of existing object vectors in order to reinforce prior successes. The crossover operation creates a trial vector \( U_{p,G+1} \) by selecting elements from the target vector \( X_{p,G} \) and the mutated donor vector \( V_{p,G+1} \). The crossover constant \( CR \) controls the probability that a trial vector parameter will come from the mutated vector \( V_{p,G+1} \), instead of from the current vector \( X_{p,G} \), and therefore ranges from 0 to 1.

**Step 6:** Check lower and upper bounds of the variables. The parameters of the child vectors must be checked for boundary conditions. If a mutated parameter exceeds some boundary constraint, one way is to select a new random but feasible value.

**Step 7:** Perform selection. To select the vectors for the next generation, each child has to be evaluated by the objective function and compared with its parent’s objective value. If the profit of the child is greater than or equal to the profit of its parent, it replaces that parent in the population, otherwise the parent will be retained in the next generation. As a result, all the individuals of the next generation are as good as or better than their counterparts in the current generation.

**Step 8:** Repeat the evolutionary cycle until \( G_{\text{max}} \) is reached.

We refer to Lampinen (2000), Lampinen & Zelinka (1999), and Fan & Lampinen (2004), for more details about the mutation schemes, values for the control parameters, and other constraint handling methods and stopping criteria.

---

4 **Computational Results**

In this section of the paper, an application of the \( M/G/c/c \) state-dependent model and the DE heuristic is presented to a system equilibrium traffic assignment problem based on an actual network. All algorithms presented in Fig. 4 and 6 were coded in C++ and are available from the authors upon request. The experiments took place on a PC, Pentium 4 3.0 GHz 2 MB CPU, 1.0 GB RAM, under Windows XP operating system.

4.1 **Algorithm Efficacy and Efficiency**

One of the main issues is to know how the algorithm results compare with earlier studies of stochastic traffic network assignment. In order to show the relationship between classical travel time estimation formulas and the new \( M/G/c/c \) formula we will consider a simple example, illustrated in Fig. 7.

![Figure 7: Two-road network and corresponding \( M/G/c/c \) model](image)

This example was considered by Prashker & Bekhor (2000), in which the two routes connecting \( A \) and \( B \) are composed by links 1 to 4, with link travel costs represented by the classical BPR formula (Bureau of Public Roads, 1964), as follows

\[
c_a(x_a) = c_a^0 \left(1 + 0.6 \left(\frac{x_a}{s_a}\right)^4\right),
\]

in which \( c_a^0 \) is the free-flow travel cost on link \( a \) and \( s_a \) is its capacity. For the experiment to be shown, we will use the free-flow on route 1-2 equal to 10 and on route 3-4, 12, as follows

\[c_1 + c_2 = 10, \ s_1 = s_2 = 1000, \ c_3 + c_4 = 12, \ s_3 + s_4 = 2000.\]

This example is also known as the two-link network, where one router is shorter (i.e., faster), but with a lower capacity, than the bypass route. Finally, to adjust the corresponding \( M/G/c/c \) model, we will use the setting presented in Tab. 2.

The results may be seen in Fig. 8, which shows the flow on route 1-2 (the shorter route) for several different values of the arrival rate, from 0 to 3,000 veh/h. Up to 500 veh/h, only the upper route carries flow, independently on the model under use. That is, we confirm here that both models should agree under light traffic loads, as claimed earlier in this paper. Significant differences start to show under heavy traffic though. In
fact, the $M/G/c/c$ model is more optimistic and allocates significant more traffic to the shorter route as the arrival rate increases. Because $M/G/c/c$ queues will reject (or block) users that arrive when the system is at capacity, the travel time will never go to infinity as in the BPR formula, leading to the $S$-shaped behavior mentioned earlier. The conclusion is that the $M/G/c/c$ based travel time formula may produce traffic assignments that make a better use of the network, once a higher part of the traffic will take the shorter (faster) route.

Another important issue here is to know how the algorithm behaves as the number of decision variables increases. We considered a sequence of experiments with the network presented in Fig. 9, composed by an arbitrary combination of links such as those presented in Tab. 3, configured in a basic split topology, in such a way that the number of decision variables considered increases. For the experiments with one decision variable, $p_{12}$, only nodes #1, #2, and #3 were considered. Notice that the decision variable $p_{13}$ is obtained by the relation $p_{12} + p_{13} = 1$. For the experiments with two decision variable, only nodes #1 to #5 were considered, and so on.

The results of the computational experiments are presented in Fig. 10. The results reported are the CPU times in seconds and the boxplots obtained from 10 runs. The running times do not increase dramatically with the number of variables, which is very encouraging and confirms the efficiency of DE algorithms for continuous optimization, but these times tend to be less predictable (the variability increases with the number of variables) and may be prohibitive in large scale networks.

4.2 A Realistic Assignment

Problem Description

One of the principal arterial road networks surrounding the Eindhoven University of Technology campus is illustrated in Fig. 11. The corresponding network representation is seen in Fig. 12, which is an interesting example because it is rather compact and also has a number of alternative routes for directing the traffic flows. This type of network model is indicative of the evacuation of a region due to a natural or man-made calamity, so the origin-destination networks with alternative routings is quite typical of this type of evacuation problem. In the experiments that follow, we would like to examine how the travel time model and the DE algorithm perform in this context.

The origin of the traffic is “o”, from the center of campus where a major parking garage is located, and the destination is “d”, the intersection of two roads. Tab. 3 illustrates the basic data necessary to implement the travel time function of the $M/G/c/c$ model and Fig. 13 depicts the queueing network representation. Notice that the capacity is determined by the speed-density curves of the particular link along with the geometry of the link. All adjusted service types are presented in Fig. 14 and all travel times as functions of the arrival rates are illustrated in Fig. 3.

Discussion

For the $M/G/c/c$ model, the capacity analysis of the routes may be determined by a bottleneck analysis of the links in the routes. There are four distinct routes, as it is shown in Table 4.

From Tab. 3, it is possible to realize that, for route #1, the bottleneck is link #7, from the point of view of capacity. Likewise, for route #2 the bottleneck is link #6, while for route #3 the bottleneck is link #4 and for route #4 the bottleneck is link #5. These bottlenecks are
Table 2: Settings for the two-road network example

<table>
<thead>
<tr>
<th>Route</th>
<th>Length(*)</th>
<th>Width (# lanes)</th>
<th>$V_1^{(*)}$</th>
<th>$V_n^{(*)}$</th>
<th>$V_i^{(*)}$</th>
<th>$c$ (veh)</th>
<th>$E[T_i]$ (h)</th>
<th>service #</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>5 (8.0)</td>
<td>1</td>
<td>30 (48)</td>
<td>27 (43)</td>
<td>12 (19)</td>
<td>1,000</td>
<td>0.1667 (10 min)</td>
<td>5</td>
</tr>
<tr>
<td>3-4</td>
<td>5 (8.0)</td>
<td>2</td>
<td>25 (40)</td>
<td>23 (37)</td>
<td>10 (16)</td>
<td>2,000</td>
<td>0.2000 (12 min)</td>
<td>6</td>
</tr>
</tbody>
</table>

(*) in miles (km); (**) in mph (km/h);

Table 3: Basic $M/G/c/c$ network data

<table>
<thead>
<tr>
<th>Link (i, j)</th>
<th>Length(*)</th>
<th>Width (# lanes)</th>
<th>$V_1^{(*)}$</th>
<th>$V_n^{(*)}$</th>
<th>$V_i^{(*)}$</th>
<th>$c$ (veh)</th>
<th>$E[T_i]$ (h)</th>
<th>service #</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (o,A)</td>
<td>0.80 (0.50)</td>
<td>5</td>
<td>25 (40)</td>
<td>23 (37)</td>
<td>10 (16)</td>
<td>800</td>
<td>0.0320</td>
<td>6</td>
</tr>
<tr>
<td>2 (A,C)</td>
<td>2.50 (1.55)</td>
<td>2</td>
<td>20 (32)</td>
<td>18 (29)</td>
<td>6 (10)</td>
<td>1000</td>
<td>0.1250</td>
<td>7</td>
</tr>
<tr>
<td>3 (A,B)</td>
<td>1.85 (1.15)</td>
<td>2</td>
<td>20 (32)</td>
<td>18 (29)</td>
<td>6 (10)</td>
<td>740</td>
<td>0.0925</td>
<td>7</td>
</tr>
<tr>
<td>4 (B,C)</td>
<td>1.00 (0.62)</td>
<td>1</td>
<td>45 (72)</td>
<td>40 (64)</td>
<td>16 (26)</td>
<td>200</td>
<td>0.0222</td>
<td>4</td>
</tr>
<tr>
<td>5 (C,B)</td>
<td>1.00 (0.62)</td>
<td>1</td>
<td>45 (72)</td>
<td>40 (64)</td>
<td>16 (26)</td>
<td>200</td>
<td>0.0222</td>
<td>4</td>
</tr>
<tr>
<td>6 (C,D)</td>
<td>0.53 (0.33)</td>
<td>2</td>
<td>30 (48)</td>
<td>27 (43)</td>
<td>12 (19)</td>
<td>212</td>
<td>0.0177</td>
<td>5</td>
</tr>
<tr>
<td>7 (B,D)</td>
<td>0.56 (0.35)</td>
<td>2</td>
<td>45 (72)</td>
<td>40 (64)</td>
<td>16 (26)</td>
<td>224</td>
<td>0.0124</td>
<td>4</td>
</tr>
<tr>
<td>8 (D,d)</td>
<td>0.62 (0.39)</td>
<td>2</td>
<td>45 (72)</td>
<td>40 (64)</td>
<td>16 (26)</td>
<td>248</td>
<td>0.0138</td>
<td>4</td>
</tr>
</tbody>
</table>

(*) in miles (km); (**) in mph (km/h);

Figure 9: Basic split topology
Table 4: Routes and lone occupant travel times

<table>
<thead>
<tr>
<th>Route</th>
<th>Links*</th>
<th>$E[T_s]$ (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 → 3 → 7 → 8</td>
<td>0.1507</td>
</tr>
<tr>
<td>2</td>
<td>1 → 2 → 6 → 8</td>
<td>0.1884</td>
</tr>
<tr>
<td>3</td>
<td>1 → 3 → 4 → 6 → 8</td>
<td>0.1782</td>
</tr>
<tr>
<td>4</td>
<td>1 → 2 → 5 → 7 → 8</td>
<td>0.2054</td>
</tr>
</tbody>
</table>

*L bottlenecks are in boldface

due to the length of the road segment and the number of lanes as well as the free flow speeds and parameters necessary for the $M/G/c/c$ model. Thus, it makes sense to route the traffic along the shortest time route which is route #1. However, due to the capacity limitations, other routes will absorb the additional overflow traffic if route #1 is saturated. Ultimately, we would like to know which routes are best given the traffic volumes that are to be accommodated.

Tab. 5 illustrates the results of applying the proposed algorithms to the traffic network. Notice that the results are sound over the different scenarios and provide interesting insights for the traffic flow behavior, if one admits $M/G/c/c$ state-dependent queueing networks as a modeling tool. Mostly, only routes #1 and #2 received flows, which makes sense. In fact, routes #3 and #4 add nodes #4 and #5 and because of that they are longer options, as we can see from column $E[T_s]$. Up to $\lambda = 2,000$ veh/h, the flow divides into routes #1 and #2. Beyond this point, it is even more advantageous to keep the whole traffic in route #1. It is curious that it was never an advantage to reroute the traffic to alternative routes and that the total traffic throughout the network even decreased when the arrival went above 2,000 veh/h, which seems to be the maximum capacity of this network. The $M/G/c/c$ state-dependent model has this well-know property of saturation indeed, as seen in previous studies (see Jain & Smith, 1997; Mitchell & Smith, 2001), caused by the rejection of blocked users, that is, users that arrived when the system was at capacity. In other words, the $M/G/c/c$ state-dependent queueing model helps to identify a capacity in terms of the maxi-
Figure 12: Traffic network representation

Figure 13: Queueing network representation

maximum amount of vehicles that can go through it per time unit.

**Sensitivity Analysis**

As part of a sensitivity analysis to see what would happen if reductions in capacity were to occur. Let us remove one lane of traffic from link 7 and reduce its speed to 25 mph. This link is a critical part of the traffic network because it is responsible for all blocking present in route #1. The results of this removal may be seen in Table 6. Again, the M/G/c/c model assigns traffic to the longer route to alleviate the congestion on the shorter routes.

One positive thing about these results is that even losing the full capacity on link #7, one of the most important, the overall throughput does not change significantly as we see different assignments only for λ = 1,000 and 2,000 veh/h. In other words, the M/G/c/c state-dependent queueing model provides robust solutions, that is, which do not change significantly with slight uncertainties and errors in the network parameters.

In summary, we believe that the M/G/c/c state-dependent models should be considered carefully as alternative ways of modeling vehicular traffic. Instead of travel time formulas that accept arrivals too close to the maximum capacity of the traffic links producing times that go to infinity, finite queueing networks will reject (or block) users under massive arrivals leading to S-shaped travel times. Without considering the blocking effect, the true capacity and the real utilization of the networks will not be revealed, as seen in this example.

**5 Conclusions and Final Remarks**

This paper has presented an overview of the traffic assignment problem in urban networks. A new heuristic algorithm was proposed to the system optimum model, which is important for a better use of the scarce resources represented by the streets and roads capacities. The M/G/c/c state-dependent model was felt to be a good fit for the problem, having generated in the past sound results under many different scenarios (Jain & Smith, 1997; Kerbache & Smith, 2000; Mitchell & Smith, 2001; Cruz & Smith, 2007). The solutions seemed to be robust as it was demonstrated by a sensitivity analysis. A case study of the application of the algorithms to a traffic assignment problem in and around the Eindhoven University of Technology campus revealed in-
interesting traffic assignment patterns under different demands and scenarios.

Finally, the authors wish to point out that the main focus of this paper was to present how the system optimum may be reached in a urban traffic network modeled as an $M/G/c/c$ state-dependent queueing system, an some new insights the approach may bring to the analysis.

There are a number of directions possible with this research. For example, a more general network may be considered, in which the number of decision variables is larger than those tested in this article. More work can be done to try to improve algorithm efficiency, especially in real life large-scale networks. Another possibility is to recognize that the $M/G/c/c$ model is also directly applicable to modeling pedestrian networks, so that many of the similar features of the travel delay function as shown in this paper apply to pedestrian dynamics.

ACKNOWLEDGMENTS

The research of Frederico Cruz has been partially funded by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) of the Ministry for Science and Technology of Brazil, grants 201046/1994-6, 301809/1996-8, 307702/2004-9, 472066/2004-8, 472877/2006-2, and 304944/2007-6, by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), grant BEX-0522/07-4, by FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais), grants CEX-289/98, CEX-855/98, TEC-875/07, and CEX-PPM-00401/08.

REFERENCES


Braess, D., Nagurney, A., Wakolbinger, T., 2005. On a paradox of traffic planning [translation from the original in German, Braess (1968)]. Transportation Science 39, 446–450.


Table 6: Traffic assignment sensitivity analysis

<table>
<thead>
<tr>
<th>( \lambda )</th>
<th>Route</th>
<th>Links</th>
<th>Assignment</th>
<th>( E[T_r] ) (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>1</td>
<td>1 → 3 → 7 → 8</td>
<td>500</td>
<td>0.1703</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1 → 2 → 6 → 8</td>
<td>0</td>
<td>0.1893</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1 → 3 → 4 → 6 → 8</td>
<td>0</td>
<td>0.1852</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1 → 2 → 5 → 7 → 8</td>
<td>0</td>
<td>0.2188</td>
</tr>
<tr>
<td>1000</td>
<td>1</td>
<td>1 → 3 → 7 → 8</td>
<td>866</td>
<td>0.1687</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1 → 2 → 6 → 8</td>
<td>134</td>
<td>0.1921</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1 → 3 → 4 → 6 → 8</td>
<td>0</td>
<td>0.1944</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1 → 2 → 5 → 7 → 8</td>
<td>0</td>
<td>0.2108</td>
</tr>
<tr>
<td>2000</td>
<td>1</td>
<td>1 → 3 → 7 → 8</td>
<td>866</td>
<td>0.1840</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1 → 2 → 6 → 8</td>
<td>134</td>
<td>0.2113</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1 → 3 → 4 → 6 → 8</td>
<td>0</td>
<td>0.2097</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1 → 2 → 5 → 7 → 8</td>
<td>0</td>
<td>0.2300</td>
</tr>
<tr>
<td>4000</td>
<td>1</td>
<td>1 → 3 → 7 → 8</td>
<td>1224</td>
<td>1.2881</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1 → 2 → 6 → 8</td>
<td>0</td>
<td>0.6859</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1 → 3 → 4 → 6 → 8</td>
<td>0</td>
<td>1.3122</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1 → 2 → 5 → 7 → 8</td>
<td>0</td>
<td>0.7062</td>
</tr>
<tr>
<td>8000</td>
<td>1</td>
<td>1 → 3 → 7 → 8</td>
<td>1223</td>
<td>1.2899</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1 → 2 → 6 → 8</td>
<td>0</td>
<td>0.8113</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1 → 3 → 4 → 6 → 8</td>
<td>1</td>
<td>1.3130</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1 → 2 → 5 → 7 → 8</td>
<td>0</td>
<td>0.8316</td>
</tr>
</tbody>
</table>


